Received October 2001
Revised November 2001
Accepted April 2002

# Measurement distortion of graphs in corporate reports: an experimental study

Vivien Beattie
*Department of Accounting, Finance and Law, University of Stirling, Stirling, UK, and*
Michael John Jones
*Cardiff Business School, Cardiff, UK*

**Keywords** *Company reports, Financial accounting, Graphs, Measurement*

**Abstract** *Graphs in corporate annual reports are a double-edged sword. While they offer the potential for improved communication of accounting information to users, the preparers of the annual reports can easily manipulate the graphs for their own interests. For over a decade, the empirical financial graphics literature has focused on examining company reporting practices. A particular concern has been measurement distortion, which violates a fundamental principle of graph construction. Unfortunately, it is not yet known whether observed levels of measurement distortion are likely to affect users' perceptions of financial performance. This study uses an experimental approach to address this issue. Pairs of graphs are shown to establish the level of difference that is just noticeable to graph readers. Six levels of "distortion" are investigated (5 per cent, 10 per cent, 20 per cent, 30 per cent, 40 per cent and 50 per cent). Results indicate that if financial graphs are to avoid distorting the perceptions of users, then no measurement distortions in excess of 10 per cent should be allowed. Users with lower levels of financial understanding appear to be most at risk of being misled by distorted graphs. Further research will be necessary to investigate whether this impact upon perceptions subsequently affects users' decisions in specific contexts.*

## Introduction

For over 200 years, graphs have been used in many technical and everyday contexts to communicate information effectively. Voluntary presentation graphics are increasingly used in the corporate annual reports of large companies in many countries (see, for example, Beattie and Jones, 2001). This increase in usage can be attributed largely to the changing role of the corporate report – from a formal, statutory document for shareholders to a major advertising and public relations document, serving multiple purposes and multiple audiences (Hanson, 1989; Squiers, 1989; Lee, 1994; Hopwood, 1996).

The communication advantages of graphs are well-established and are fourfold. First, graphs attract our attention, especially if their visual saliency is increased by the use of colour (Leivian, 1980). Such visual representations become "graphical sound bites" (Henry, 1995, p. 35). Second, because graphs rely on spatial, rather than linguistic intelligence, we can use our dominant visual sense to "see" the data in a direct and immediate way. This facilitates comparisons and the identification of patterns, trends and anomalies (Korol, 1986; Harris, 1996, p. 164). Third, the data can be readily retreived (Wainer,

Accounting, Auditing & Accountability Journal,
Vol. 15 No. 4, 2002, pp. 546-564.

DOI 10.1108/09513570210440595

1992). Fourth, in the specific context of corporate annual reports, graphs provide oases of colour and interest that enliven the presentation of information.

Unfortunately, it is also well-established that the preparers of corporate reports have incentives to manipulate the content of these reports, or at least to manage the impression conveyed by them. This behaviour is generally aimed at creating a more favourable view of the company's performance than is warranted. For studies that discuss the impact of these incentives on accounting disclosure choices, see Watts and Zimmerman (1986) (positive accounting theory); Tweedie and Whittington (1990) and Revsine (1991) (the "selective financial misrepresentation" hypothesis); Murphy and Zimmerman (1993) (the "cover-up" hypothesis); Lewellen *et al.* (1996) (self-serving behaviour); and Preston *et al.* (1996, p. 119) (impression management). This manipulation can take a number of forms, including biasing (the selection of favourable information items) and focusing (the enhancement of degradation of aspects of the information set) (Birnberg *et al.*, 1983).

In recent years, regulators have paid increasing attention to those aspects of the annual report package that lie outside the audited financial statements, recognising the importance of "financial communication rather than mere financial reporting" (FRC, 1995, p. 23, 1999). These aspects include narratives, graphs and photographs. Regulators have also shown increasing concern regarding the potential for this discretionary material to be manipulated (this concern follows naturally from a period of intense concern regarding the manipulation of the financial statement numbers themselves (e.g. Levitt, 1998)).

For example, in the UK, the Department of Trade and Industry's (DTI) review of company law has proposed that the operating and financial review should be "reviewed" by auditors (DTI, 2000, pp. 189-90), a proposal that goes far beyond the current UK auditing requirements in relation to information in documents containing audited financial statements (APB, 1999). Moreover, the UK's Accounting Standards Board (ASB) has issued a discussion paper that recognises that graphs are a powerful medium of communication and makes five recommendations regarding the use of graphs in annual reports. These recommendations cover selectivity in the graphs shown, selectivity in the length of time series shown, measurement distortion, the need for simple two-dimensional formats and the need for related commentary to be located adjacent to the graphs (ASB, 2000, pp. 28-9). Similarly, the sustainability reporting guidelines issued by the Global Reporting Initiative (GRI, 2000, pp. 8-9) comment on the value of graphs in reports, but note the importance of neutrality in presentation.

Beattie and Jones (1992a, p. 1) identify three forms of graphical infidelity. Selectivity relates to bias regarding the choice of variables graphed. Measurement distortion occurs where the physical representation of the numbers on the graph is not directly proportionate to the underlying numbers. Finally, presentational enhancement arises where the design of the graph in some way enhances or degrades certain aspects of the information set, for example because of the use of three-dimensional forms or because the final year's results are "highlighted" in a brighter colour.

Measurement distortion is the topic of this particular paper. Prior empirical studies of financial graphs have found that material measurement distortions (usually defined by the prior literature as distortions in excess of 5 per cent) occur in a significant number of graphs and that these distortions generally give a more favourable view of the companies' performance than is warranted.

Unfortunately, although the incidence of measurement distortions has been empirically demonstrated on many occasions (see, for example, Steinbart, 1989; Mather *et al.*, 2000), these studies have been conducted without any insight into what level of measurement distortion is generally sufficient to influence a user's perception of a company's performance. As a result, the practical significance of the findings from empirical studies of company practice remains unclear.

The purpose of the present study is to address this issue. For the first time, we provide experimental evidence to ascertain what level of measurement distortion is sufficient to affect the perceptions of users. Our results not only frame the prior financial graphics research, they also provide much-needed guidance for future research in this area.

The remainder of this paper is structured as follows. Section two reviews the theoretical and empirical literature relevant to measurement distortion and formally presents the research question and hypothesis. Section three outlines the experimental method adopted. The results are then presented in section four. A discussion section then uses our results to contextualise the prior research; it also discusses the limitations of the study and identifies directions for further research. The final section concludes.

## Prior literature

The volume of research into financial graphs has increased rapidly since Steinbart's seminal paper on US data, published in 1989. Researchers in Australia, the UK, the USA and Hong Kong have investigated the use and abuse of graphs in annual reports and prospectuses in both national and cross-national studies. They have documented instances of selectivity, measurement distortion and presentational enhancement. A key aspect of most of these studies has been the empirical documentation of measurement distortion. At least nine studies (see Table I) have investigated this issue in detail.

Studies of measurement distortion in graphs have all used a graph discrepancy index. This index is a way of measuring the misrepresentation of the underlying numerical data when they are graphically portrayed. The graph discrepancy index used in financial accounting studies originates from Tufte (1983). Tufte is a famous graphical researcher who devised a "lie factor". In the accounting literature, Taylor and Anderson (1986) adapted Tufte's lie factor as follows:

**Table I.** Key features and findings of studies of measurement distortion in corporate reports

| Study | Country | Companies studied | Mean level of measurement distortion[a,b] | Incidence of material measurement distortion | Favourable/unfavourable; association with corporate performance |
|---|---|---|---|---|---|
| Johnson *et al.* (1980) | USA | 423 graphs from 50 *Fortune* 500 annual reports for 1977 and 1978 | Not given | Not given | Not given |
| Steinbart (1989) | USA | 319 *Fortune* 500 annual reports for 1986 | Mean = +11% (across 3 KFVs) | 26% (>\|10%\|) | Significant |
| Beattie and Jones (1992b) | UK | 1989 annual reports of 240 large companies | Mean = +10.7% (across 4 KFVs) | 30% (>\|5%\|) | Not significant |
| Mather *et al.* (1996) | Australia | 1991 and 1992 annual reports for (i)143 top listed companies and (ii) 44 not-for-profit entities | (i) Mean = +16.4% (across 4 KFVs)<br>(ii) Mean = +105.6% (across max. of 3 graphs per entity) | (i) 13.3% (>\|10%\|)<br>(ii) 51% (>\|5%\|) | (i) Significant (where performance measured as change in variable over period graphed)<br>(ii) Not applicable |
| Beattie and Jones (1997) | USA/UK | 1990 annual reports of 85 US and 91 UK leading companies | US = +16%<br>UK = +7%<br>(across 4 KFVs) | US: 24% (>\|5%\|)<br>UK: 24% (>\|5%\|) | Not given |
| Frownfelter and Fulkerson (1998) | 12 countries | 74 companies covering 12 countries | Mean absolute GDI of 136% | Overall 68% (>\|5%\|) | Significant |
| Beattie and Jones (1999) | Australia | 1991 annual reports of 89 leading listed companies | +3.5% (adjusted to take into account favourable and unfavourable trends) (across 4 KFVs) | 34% (>\|5%\|) | Not significant |
| Mather *et al.* (2000) | Australia | 484 IPO prospectuses issued pre 1994 | +86% (across 200 non-KFVs)<br>+2% (across 4 KFVs) | Non-KFVs: 63%<br>KFVs: 48%<br>(>\|5%\|) | KFVs: unfavourable distortion significantly more likely than favourable distortion |

*(continued)*

| Study | Country | Companies studied | Mean level of measurement distortion[a,b] | Incidence of material measurement distortion | Favourable/unfavourable; association with corporate performance |
|---|---|---|---|---|---|
| Beattie and Jones (2000) | 6 countries: Australia, France, Germany, The Netherlands, UK and USA | 1991 or 1992 annual reports of 50 of the top 100 companies in each country | Australia: +1%<br>France: +36%<br>Germany: –13%<br>The Netherlands: +3%<br>UK: +86%<br>USA: +30% | Not given | Not given |

**Notes:**
[a] KFV = key financial variable, typically sales, profits, earnings per share and dividends per share
[b] GDI = graph discrepancy index.
[c] Courtis (1997) reports on a study of the 1994-95 annual reports of 327 companies listed on the Hong Kong Stock Exchange, 35 per cent of which include graphs. Of these graphs, 52 per cent were classified as "misleading", in that they violated one of five graph construction guidelines (one of which related to measurement distortion). No specific figures are reported in the paper, however

**Table I.**

$$\text{Graph discrepancy index} = [(a/b) - 1] \times 100\%$$

where

$$a = \text{percentage change (in cms) depicted in graph, i.e.} \quad \frac{\text{height of last column} - \text{height of first column}}{\text{height of first column}} \times 100\%$$

$$b = \text{percentage change in data.}$$

If, for example, a company's sales rise from £5m to £10m over a five year period, and this is portrayed in a column graph with the height of the column in year 1 being 5cm and the height of the year 5 column being 10.5cm, then the graph discrepancy index is +10%:

$$\text{GDI} - [(110/100) - 1] \times 100\%$$

where

$$\begin{aligned} a &= [(10.5 - 5)/5] \times 100\% \\ &= 110 \\ b &= [(10 - 5)/5] \times 100\% \\ &= 100 \end{aligned}$$

In the absence of measurement distortion, the value of the index is zero. Positive (negative) values indicate the per centage by which the trend in the data is exaggerated (understated) by the graph. Non-zero values can arise from either specific features of graph design (such as a non-zero or a broken axis) or from inaccurate draughtsmanship. To determine whether the distortion is favourable (i.e. flattering) or unfavourable, the nature of the variable and the direction of the trend line must be taken into account. The four most commonly graphed financial variables are sales, profits, earnings per share and dividends per share (Beattie and Jones, 1992a,b). These variables are often termed key financial variables (KFVs). In the case of these variables, higher values are seen as "better" than lower values. Consequently, the exaggeration of an upward trend and the understatement of a declining trend both give a more favourable impression of a company's performance. In annual reports, it has in recent years been more usual to encounter upward, rather than downward, trends in KFVs.

There is, however, no empirical evidence regarding what constitutes a material distortion. Some authors have, however, speculated. For example, Tufte (1983, p. 57) argues that distortions in excess of 5 per cent indicate "substantial distortion, far beyond minor inaccuracies in plotting".

In the accounting literature, many studies systematically document the level of measurement distortion found in corporate annual reports and prospectuses, most studies dealing with corporate annual reports. A few studies also investigate the features of graph design (such as a non-zero axis) that give rise to the measurement distortion. Typically, these studies distinguish between favourable and unfavourable measurement distortion and investigate the association between measurement distortion and measures of company performance. The majority of studies are single-country. However, a few more recent studies undertake cross-national comparisons. The key features and findings of these studies in relation to measurement distortion are summarised in Table I.

Table I demonstrates a widespread incidence of material distortion. Using a 5 per cent cutoff, the percentage of distorted graphs ranges from 24 per cent (Beattie and Jones, 1997) to 68 per cent (Frownfelter and Fulkerson, 1998). The mean level of measurement distortion found ranges from –13 per cent (in Germany: Beattie and Jones, 2000) to +86 per cent (in the UK: Beattie and Jones, 2000). A few studies also report the frequency distribution of graph discrepancy scores. For example, Beattie and Jones (1992b, p. 300: not shown in table) report that 70 per cent of graphs show no material absolute distortion (i.e. $<|5$ per cent$|$), 10 per cent are distorted by between $|5$ per cent$|$ and $|10$ per cent$|$; a further 10 per cent are distorted by between $|10$ per cent$|$ and $|25$ per cent$|$, 4 per cent are distorted by between $|25$ per cent$|$ and $|50$ per cent$|$ and 7 per cent are distorted by more than $|50$ per cent$|$. The results from a more extensive six-country study which reports frequency distributions are shown in Table II. This Table shows that 26 per cent of key financial graphs are distorted by between $|5$ per cent$|$ and $|50$ per cent$|$, with only 9 per cent being distorted by more than $|50$ per cent$|$.

These prior data thus show high levels of measurement distortion. However, the critical question remains unanswered. What level of distortion triggers a change in users' perception of corporate performance? The answer would contextualise the evidence from prior studies. Beattie and Jones (1999, p. 59) note the "need for experimental testing of this threshold in the context of financial graphs".

Somewhat surprisingly, there have been almost no published experimental studies of the impact of measurement distortion in financial graphs on the

**Table II.** Frequency distribution of observed graph discrepancy scores across six countries

| Graph discrepancy score ($x$) | Australia % | France % | Germany % | The Netherlands % | UK % | USA % | Mean % |
|---|---|---|---|---|---|---|---|
| $x < \|5\%\|$ | 62 | 61 | 57 | 75 | 69 | 66 | 65 |
| $\|5\%\| < x < \|10\%\|$ | 10 | 12 | 22 | 9 | 12 | 13 | 13 |
| $\|10\%\| < x < \|25\%\|$ | 13 | 11 | – | 8 | 9 | 8 | 8 |
| $\|25\%\| < x < \|50\%\|$ | 10 | 6 | – | 4 | 3 | 5 | 5 |
| $x > \|50\%\|$ | 5 | 10 | 21 | 4 | 7 | 8 | 9 |
| Total | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

**Source:** Extracted from Beattie and Jones (1996) perceptions of users. Indeed, the only such study of which we are aware is Taylor and Anderson (1986). In their study, measurement distortion is merely one of several graphical features investigated[1]. Taylor and Anderson (1986) create seven pairs of graphs. In each case, the second graph incorporated some kind of distortion. They showed one of each pair of graphs to a group (number unspecified) of commercial loan officers. The officers were then asked for their perception of company performance. One of the seven graph pairs addressed the issue of measurement distortion, by introducing a non-zero axis, producing a graph discrepancy index of a staggering +884 per cent. Not surprisingly, an exaggeration of this magnitude resulted in "a misleadingly favourable impression of company performance" (p. 127).

Statistical graphics researchers have produced, based on psychological theory and experiments, a theory of graph comprehension (i.e. graphical perception and visual processing)[2]. They have developed a specialist theory of graphical perception (i.e. the visual decoding of a graph's quantitative information) using the theory of visual information processing (Cleveland and McGill, 1987; Kosslyn, 1989, 1994). It is argued that, in reading a graph, we initially perform rapid visual scans to detect the geometric patterns that form the basis of our inferences about the data's behavior. These initial perceptual tasks may (but, importantly, may not) be followed by more highly cognitive tasks such as scale reading. This theory of graphical perception provides the basis for the present study.

### *Research question and hypothesis*

The research question addressed in this study is as follows: what is the level of measurement distortion that would trigger a change in the user's perception of a company's performance?[3] Importantly, this question assumes that there is some (unknown) level at which the two concepts are causally related. Formally, we test the null hypothesis that:

*H0*. A change in the underlying trend line shown in a five-year column graph does not affect the perception of company performance.

The prior literature speculates that |5 per cent| distortion is the minimum level that could influence perceptions. Meanwhile, empirical studies of the levels observed in practice suggest that very few graphs display distortion in excess of |50 per cent|. We, therefore, test six specific levels of graphical measurement distortion ranging from 5 per cent to 50 per cent. (i.e. 5 per cent, 10 per cent, 20 per cent, 30 per cent, 40 per cent, 50 per cent).

## Methods

### *Subjects*

A total of 52 second year business studies students who had completed a one-year course in accounting participated in the experiment[4]. In experimental studies, the choice of subject is critical. In the present study, students were appropriate subjects because the study elicited general perceptions of company performance rather than a context-specific decision or judgment. In other words, the experimental task focused on general visual processing skills, rather than the use of a specific body of knowledge learned in the real world. In such circumstances, the use of experts is not necessary (and may indeed be undesirable as the task is unlikely to be "representative"). Indeed, it is now recommended that experimenters avoid using professional subjects under these conditions, because it imposes an unnecessary burden on researchers' time and funding sources and because of the negative externalities involved (i.e. other researchers may find it more difficult to gain access to this valuable resource) (Libby *et al.*, 2001, p. 54)[5].

*Stimuli*

The experimental stimuli consisted of five-year, time series column graphs of the type most commonly found in annual reports. The graphic displays, and the experimental task, were intentionally kept abstract and simple, in order to focus on the impression conveyed from a perceptual analysis of the graph's visual components. This simple graphical decoding task is the predominant form of processing used when viewing annual report graphs. Typically, annual report graphs have neither a *y*-axis nor axis labels (see, for example, Beattie and Jones' (1992a, pp. 26-7) findings for 154 key financial graphs). Moreover, scale values were omitted to discourage highly cognitive processing, following Simcox (1984). All graphs were coloured black, to avoid any confounding perceptual effects relating to colour[6].

An initial base graph was constructed showing a steady increase of 100 per cent over the five-year period (i.e. a straight implicit trend line). This is shown as graph X in Figure 1. A further set of six graphs was then constructed, each showing one of six levels of the independent variable (5 per cent, 10 per cent, 20 per cent, 30 per cent, 40 per cent and 50 per cent). These levels are different enough for the experiment to have sufficient power to yield strong effects, yet are representative of the range observed in practice (Libby *et al.*, 2001, p. 44). The measurement distortion was incorporated throughout the graph, such that the implicit trend line remained straight, but at a steeper angle.

A major issue in developing the experimental stimuli was how to operationalise the independent variable (measurement distortion) without introducing demand effects (i.e. alerting the subjects to the specific nature of the experiment). In order to minimise this potential problem, each experimental trial was presented as a comparison task. Each distorted graph was paired with the undistorted base graph, with both orders of presentation being included, giving rise to 12 trials. The graphs were labelled X and Y. For an example, see Figure 1. The data area of the undistorted graph (graph X) measured 10cms tall by 5.8cm wide. The first specifier was 5cm high and the last specifier was 10cm. The specifiers were 0.8cm wide and the four interspaces were 0.45cm. By contrast, the distorted graph (graph Y) was 10.5cm tall by 5.8cm wide. The first specifier was 5cm, however, the final specifier was 10.5cm. The specifier width and interspaces were the same as for graph X. Thus, graph Y portrays a 10 per

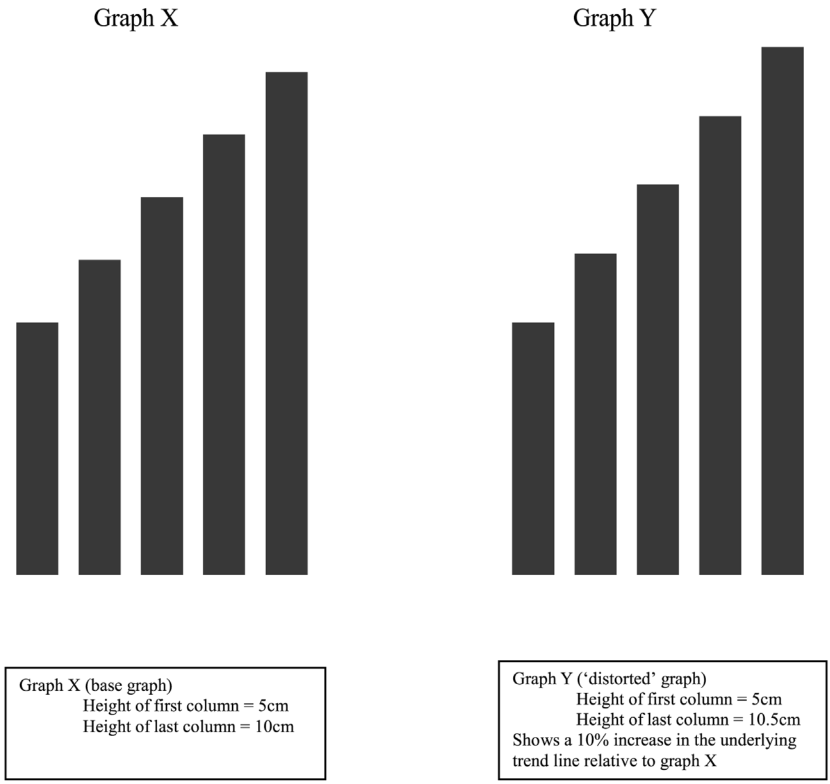


**Figure 1.** Example of experimental materials used

cent increase in the underlying trend line compared to graph X (for an illustration of graphic components, see Beattie and Jones (1998, p. 7)).

*Procedure*

Subjects were informed that the purpose of the experiment was to examine the effect, if any, of differences in graph structure on the perceptions of the data portrayed. They were also told that all the graphs to be shown represented time series data of earnings per share, which is one of four key financial variables frequently reported graphically by companies. An example of the graphic displays was shown at the start of the experiment. The instructions to subjects were given in writing and were worded as neutrally as possible to reduce the likelihood of subjects guessing the specific experimental hypothesis and answering in line with the researchers' expectations.

On a given trial, the graphic display was projected onto an overhead screen for a fixed duration of three seconds. This short display time was selected, following Cleveland and McGill (1987), in order to prevent highly cognitive processing. Subjects were shown each of the 12 graphical stimuli in random order sequence. They were asked to indicate (by circling the appropriate response on a data collection sheet) whether they noticed a difference between the performance portrayed in each graph. The three response categories were "X more favourable than Y", "no difference", and "Y more favourable than X", coded 1 to 3 respectively. The graph showing the larger data increase (i.e. the graph incorporating measurement distortion) was, beyond a certain perception threshold, expected to be perceived as showing the more favourable financial performance.

In a debriefing following the experiment, subjects were asked to indicate the level of confidence (CONF) in their responses on a five-point Likert scale (1 = not at all confident to 5 = very confident). A positive relationship is expected. In addition, subjects were asked to provide background details regarding individual difference variables that could affect their ability to detect differences (Libby and Lewis, 1977). These background variables were format preference (PREF); graph-reading ability (i.e. graphical literacy) (ABILITY); level of financial understanding (FINUND); and gender (GENDER). PREF was measured by asking subjects to indicate their personal preference for one of four options: text, tables, graphs, no preference. A dummy variable was created that took the value of one if the preferred format was graphs and the value of zero otherwise. ABILITY and FINUND were self-rated using five-point Likert scales (1 = very poor to 5 = very good) [7]. A dummy variable for GENDER took the value of one for "female" and zero otherwise.

To capture each subject's overall accuracy in detecting differences between the graphs, an accuracy measure was calculated. Four variants were used (AccA to AccD) and the detailed scoring rules are outlined in Table III. Two variants (AccB and AccD) incorporate the nature of the individual's error, i.e. the situation where no difference was perceived when, in fact, there was one,

**Table III.**
Accuracy measures and scoring rules

| Accuracy measure | Scoring rules | Range | Mean | Std dev. |
|---|---|---|---|---|
| AccA | 4 points if correct; 0 otherwise | 0-48 | 31.0 | 5.76 |
| AccB | 4 points if correct; 2 points if respond "no difference"; 0 otherwise | 0-48 | 38.8 | 3.27 |
| AccC | 8 points if correct and GDI = 5%;<br>6 points if correct and GDI = 10%;<br>4 points if correct and GDI = 20%;<br>3 points if correct and GDI = 30%;<br>2 points if correct and GDI = 40%;<br>1 point if correct and GDI = 50%; 0 otherwise | 0-48 | 21.7 | 5.95 |
| AccD | 8 points if correct and GDI = 5%;<br>6 points if correct and GDI = 10%;<br>4 points if correct and GDI = 20%;<br>3 points if correct and GDI = 30%;<br>2 points if correct and GDI = 40%;<br>1 point if correct and GDI = 50%;<br>half points if respond "no difference"; 0 otherwise | 0-48 | 34.0 | 3.78 |

compared to the situation where the existence of a difference was perceived but in the incorrect direction. The latter is viewed as a more serious error. Two variants (AccC and AccD) take into account the magnitude of the measurement distortion, on the basis that larger differences are easier to detect.

Points are summed across the trials to give an overall accuracy measure. The variants are constructed to produce identical ranges (i.e. 0 to 48). Variant AccA is the simplest measure, while variant AccD is the most sophisticated, as it takes into account both the nature of the error and the magnitude of the measurement distortion. The correlations between the accuracy measures are all very high and are all highly significant ($p < 0.0001$). The lowest correlation is that between AccA and AccD ($r = 0.772$).

The following model relating graph judgment accuracy to individual difference variables was estimated using OLS regression:

$$\text{Accuracy} = \text{f}(\text{PREF},\ \text{ABILITY},\ \text{FINUND},\ \text{GENDER})$$

Positive relationships are expected in all cases except GENDER, where the direction of effect is unclear. In relation to FINUND, since the graphs relate to a financial variable (EPS), those with a lower level of financial understanding may be less able to interpret the performance portrayed in the graph, irrespective of their level of graphical literacy. The only significant correlations between the background variables were significant negative correlations between ABILITY and GENDER, and between FINUND and GENDER, indicating that males generally rate both their graphical literacy and their financial understanding as being higher than females.

## Results

The main results are presented in Table IV, which gives the frequency distribution of responses, summary statistics and the results of a series of *t*-tests to investigate whether the mean response to each trial is different from 2 (no difference). The trials have been reordered for presentation purposes, and are shown in ascending order of magnitude of the dependent variable, first for the six trials where graph X was the base graph (panel A) and then for the six trials where graph Y was the base graph (panel B). In panel A, the mean responses range from 2.04 for a 5 per cent difference between the graphs (close to the value 2: no difference detected) to 2.96 for a 50 per cent difference between the graphs (close to 3: Y more favourable than X). In panel B, the mean responses range from 1.85 for a 5 per cent difference between the graphs (close to the value 2: no difference detected) down to 1.02 for a 50 per cent difference between the graphs (close to 1: X more favourable than Y). For each trial, the mean response is in the expected direction. It is also clear from a visual inspection of the data that, as the level of “distortion” rises, so does the proportion of subjects that correctly identified a difference between the two graphs (i.e. the further the mean score departs from 2).

It is apparent, however, that at low levels of “distortion” (5 per cent and 10 per cent), the majority of subjects perceive no difference between the two

**Table IV.**
Descriptive statistics for the 12 trials

| Trial no. | Trial details [a] | No. in response category: X more fav. Y (coded 1) | No. in response category: No diff. (coded 2) | No. in response category: Y more fav. X (coded 3) | Summary statistics: Mean | Summary statistics: Std dev. | Summary statistics: Prob. (mean = 2) [c] |
|---|---|---|---|---|---|---|---|
| Panel A | | | | | | | |
| 5 | X; Y(5%) | 1 | 48 | 3 | 2.04 | 0.28 | 0.3220 |
| 6[b] | X; Y(10%) | 1 | 38 | 12 | 2.22 | 0.55 | 0.0276** |
| 3 | X; Y(20%) | 2 | 21 | 29 | 2.52 | 0.58 | <0.0001*** |
| 8 | X; Y(30%) | 1 | 7 | 44 | 2.83 | 0.43 | <0.0001*** |
| 7 | X; Y(40%) | – | 3 | 49 | 2.94 | 0.24 | <0.0001*** |
| 2 | X; Y(50%) | – | 2 | 50 | 2.96 | 0.19 | <0.0001*** |
| *Panel B* | | | | | | | |
| 11 | X(5%); Y | 10 | 40 | 2 | 1.85 | 0.46 | 0.0194** |
| 1 | X(10%); Y | 17 | 30 | 5 | 1.77 | 0.61 | 0.0091*** |
| 4 | X(20%); Y | 39 | 11 | 2 | 1.29 | 0.54 | <0.0001*** |
| 9 | X(30%); Y | 50 | 2 | – | 1.04 | 0.19 | <0.0001*** |
| 12 | X(40%); Y | 48 | 3 | 1 | 1.10 | 0.36 | <0.0001*** |
| 10 | X(50%); Y | 51 | 1 | – | 1.02 | 0.14 | <0.0001*** |

**Notes:**
[a] The figure in parentheses shows the amount by which the graph is "distorted"; the trials are shown in ascending order of magnitude of the independent variable, first for the six trials where graph X was the base graph and then the six trials where graph Y was the base graph. Note that the closer the mean score is to 2, the poorer is the user's ability to perceive the difference between the graphs
[b] One subject did not respond to this trial.
[c] ** Significant at the 5 per cent level; *** significant at the 1 per cent level.

graphs (i.e. less that 50 per cent are able to discern differences of this magnitude). In 9 out of 12 trials, the mean response is significantly different from 2 (no difference) at the 0.01 level. These trials relate to the higher levels of distortion – 50 per cent, 40 per cent, 30 per cent, 20 per cent and one of the 10 per cent trials. At the 20 per cent level of "distortion", between 50 per cent and 75 per cent of subjects perceive a difference and at levels of 30 per cent and above, the proportion rises to 85 per cent and higher.

In the remaining three trials, the mean response is not statistically different from 2 (no difference) at the 0.01 level. In fact, the mean response for trial 5, which portrays 5 per cent distortion, is not significantly different from 2 at any conventional level of significance. The situation for trials 6 and 11 (which portray 10 per cent and 5 per cent distortion respectively) is less clear cut, as there is significance at the 0.05 level but not at the 0.01 level. Thus, *H0* is accepted at the 0.01 level in respect of the 5 per cent level of measurement distortion and rejected in respect of levels of measurement distortion of 20 per cent or higher. The conclusion with respect to the 10 per cent level of measurement distortion is less certain.

Overall, therefore, these results suggest that the vast majority of users would not notice a 5 per cent level of measurement distortion whereas a 20 per cent level and above would be noticed. At the 10 per cent level, the evidence is more mixed. In both trials (1 and 6), a majority of subjects perceive no difference, however, the mean response is statistically different from 2 at the 0.01 and 0.05 levels respectively. Given this evidence, it appears prudent to suggest that, if financial graphs are to avoid distorting the perceptions of users, then no measurement distortions in excess of 10 per cent should be allowed.

*Impact of individual difference variables*

The correlation between level of confidence (CONF) and accuracy was, as expected, positive, but not significant for any variant of the accuracy measure. The highest correlation was with AccB ($r = 0.118$, $p = 0.40$).

The model relating accuracy to individual difference variables was estimated for each of the four accuracy measures (see Table V). The best fit is obtained using the most sophisticated accuracy measure, AccD. Although the model overall is significant at the 10 per cent level, the adjusted $R^2$ is only 9 per cent. In general, the coefficient estimates have the expected sign (the exception being the format preference dummy variable for the AccD variant). The constant term is highly significant in each model, indicating the baseline accuracy level to be expected. Only the FINUND variable is significant (at either the 0.01 or 0.05 level) across all variants. None of the other three variables is significant using any variant. Thus, it appears that higher levels of declared financial understanding are associated with greater accuracy in perceiving differences in corporate performance that are portrayed graphically.

The model was re-estimated using only the six trials involving the three smaller levels of difference (i.e. 5 per cent, 10 per cent and 20 per cent). At levels of difference higher than this, 85 per cent or more of subjects correctly

**Table V.**
OLS regression results using judgment accuracy as the dependent variable

| Independent variable | Dependent variable AccA | AccB | AccC | AccD |
|---|---|---|---|---|
| | Coefficient estimates | | | |
| Constant | 21.52*** | 33.38*** | 10.88*** | 27.01*** |
| PREF | 1.03 | 0.15 | 0.37 | –0.26 |
| ABILITY | 1.01 | 0.57 | 0.86 | 0.67 |
| FINUND | 1.77** | 1.13** | 2.21** | 1.48*** |
| GENDER | 0.48 | 0.17 | 2.10 | 0.88 |
| Adj $R^2$ | 0.06 | 0.08 | 0.07 | 0.09 |
| $F$-statistic | 1.86 | 2.18 | 1.94 | 2.32 |
| $p$-value | 0.13 | 0.08* | 0.12 | 0.07* |

**Note:** * significant at the 10 per cent level; ** significant at the 5 per cent level; *** significant at the 1 per cent level

identified the difference between the two graphs, and so there was little variation in the responses. This restriction improved the overall fit of the model in all cases and FINUND was significant at the 0.01 level in all cases. However, no other variable approached significance.

## Discussion

Our results suggest that, as a rule of thumb, measurement distortions in excess of 10 per cent may cause users' perceptions of financial graphs' trendlines to be altered. If we compare this materiality threshold of 10 per cent with the findings of prior research into corporate practices, we can, for the first time, interpret the practical significance of their findings in a meaningful way.

Of the nine studies summarised in Table I, eight provide information about the mean level of measurement distortion. In each of these studies (with the exception of Beattie and Jones (1999)), a mean measurement distortion of greater than |10 per cent| is reported for at least one of the sub-groups of companies studied. These groups cover several countries: Australia (Mather *et al.*, 1996); France (Beattie and Jones, 2000); Germany (Beattie and Jones, 2000); the UK (Beattie and Jones, 1992b, 2000); and the USA (Steinbart, 1989; Beattie and Jones, 1997, 2000). In four cases, the mean measurement distortion is greater than |50 per cent| for at least one of the sub-groups of companies studied.

Table I also provided evidence on the overall incidence of material measurement distortion, although only two studies use the 10 per cent materiality threshold. Steinbart reports that 26 per cent of US key financial graphs had measurement distortion in excess of |10 per cent|, while Mather *et al.* (1996) report that 13.3 per cent of top listed Australian companies displayed distortions of this magnitude. A detailed frequency distribution of measurement distortion in key financial graphs contained in 50 corporate annual reports from top companies in each of six countries is provided by Beattie and Jones (1996) – see Table II. This shows that 28 per cent of Australian graphs, 27 per cent of French graphs, 21 per cent of German graphs, 16 per cent of Dutch graphs, 19 per cent of UK graphs and 21 per cent of US graphs had measurement distortion of greater than |10 per cent|. It is also worth emphasising that, when related to corporate performance, these distortions were mostly favourable (i.e. they showed company performance in a more favourable light that was warranted by the underlying data).

Our research suggests strongly that a 10 per cent threshold is appropriate when assessing the practical significance of observed levels of measurement distortion. Given the extant evidence that a significant proportion of financial graphs are distorted (often favourably) by more than this amount, it can be concluded that, in many cases, users obtain a more favourable impression of the company's performance than is warranted.

The principal limitations of this study are that only one graphic form is used and that only one level of data change (i.e. an increase of 100 per cent over five years) is investigated. Future research should explore the impact of: varying both the design and construction characteristics of the basic graphic form (for example, using concave/convex trend lines; adding value labels; and/or varying the overall proportions of the data area (i.e. the shape parameter). It should also investigate the impact of levels of data change other than that used in this study. The nature and complexity of the task set could also be varied.

A more general limitation derives from the study's generic setting. Perceptions of company performance are not, therefore, related to context-specific decision situations. Thus, no strong inferences can be drawn regarding the level of graph distortion that would influence users' decisions and/or judgments in specific decision contexts.

## Summary and conclusions

Prior research into corporate graphical reporting practices has shown that, across a number of studies, there are substantial levels of measurement distortion. For example, a wide-ranging six-country study found that 26 per cent of key financial graphs are distorted by between |5 per cent| and |50 per cent|, with 9 per cent being distorted by more than |50 per cent|. However, to date, there has been no attempt to assess, empirically, what constitutes material measurement distortion, i.e. what change in the underlying trend line of a column graph is necessary to change the user's perception of company performance. This study investigates this issue. Drawing upon visual information processing theory, an experimental task was constructed that required subjects to compare the performance of two companies based on graphical stimuli. The graphs were constructed so that the performance portrayed of one of the companies was better than the other (the levels of difference being 5 per cent, 10 per cent, 20 per cent, 30 per cent, 40 per cent and 50 per cent) thus capturing the impact of measurement distortion.

The results suggest that, to avoid distorting the perceptions of users, no measurement distortions in excess of 10 per cent should be allowed in financial graphs. This evidence is of value to both the statistical graphics and the accounting disciplines. In particular, it provides an empirical and justifiable basis for policy-makers to make explicit recommendations about measurement distortion. Users with lower levels of financial understanding appear to be most at risk of being misled by distorted graphs.

This study has shown that users' perceptions of company performance are altered at levels of measurement distortion of approximately 10 per cent. Whether or not these altered perceptions carry through to affect specific decisions remains a matter for further research.

## Notes

1. Experimental research currently represents only a small part of empirical financial accounting research, relative to the dominant archival method.
2. A detailed review of this literature is beyond the scope of this paper and only the relevant issues are covered here (the interested reader is referred to Beattie and Jones (1992a, chapter 2) and references therein).
3. This research question exhibits the four characteristics important to effective experimental financial accounting research: it has external validity, in that it captures important aspects of the target environment; it draws upon theory in the fundamental discipline of psychology and contributes to both financial accounting and to psychology; the relevant theory is stated in the most general terms possible; and the research question is based on a theory that describes causal relationships between concepts (Libby *et al.*, 2001, pp. 39-41).
4. The experiment was conducted at Cardiff Business School in the UK during the mid-1990s. The data are not time-sensitive since the task concerns general visual processing skills.
5. A recent study that uses student subjects under similar circumstances is Maines and Hand (1996).
6. Libby *et al.* (2001, p. 43) argue that low stimulus realism is appropriate in experiments that test very general psychological theories (such as the visual information processing theory being tested here).
7. The ABILITY and FINUND variables were self-rated and may well contain measurement error. However, there is no a priori reason to believe that self-rating is likely to introduce systematic bias. We chose not to make use of other (more elaborate but not necessarily superior) possible ways of developing a proxy for the level of financial understanding.

## References

APB (1999), "Other information in documents containing audited financial statements", *Statement of Auditing Standard 160*, Auditing Practices Board, London.

ASB (2000), "Year-end financial reports: improving communication", discussion paper, February, Accounting Standards Board, London.

Beattie, V.A. and Jones, M.J. (1992a), "The communication of information using graphs in corporate annual reports", *Certified Research Report 31*, Chartered Association of Certified Accountants, Certified Accountants Educational Trust, London.

Beattie, V.A. and Jones, M.J. (1992b), "The use and abuse of graphs in annual reports: a theoretical framework and empirical study", *Accounting and Business Research*, Autumn, pp. 291-303.

Beattie, V.A. and Jones, M.J. (1996), *Financial Graphs in Corporate Annual Reports: A Review of Practice in Six Countries*, Research Report, Institute of Chartered Accountants in England and Wales, London.

Beattie, V.A. and Jones, M.J. (1997), "A comparative study of the use of financial graphs in the corporate annual reports of major US and UK companies", *Journal of International Financial Management and Accounting*, Vol. 8 No. 1, pp. 33-68.

Beattie, V.A. and Jones, M.J. (1998), "Graphical reporting choices: communication or manipulation?", *Research Report 56*, Chartered Association of Certified Accountants, Certified Accountants Educational Trust, London.

Beattie, V.A. and Jones, M.J., (1999), "Do Australian graphs give a true and fair view?", *Abacus*, Vol. 35 No. 1, pp. 46-76.

Beattie, V.A. and Jones, M.J. (2000), "Impression management: the case of inter-country financial graphs", *Journal of International Accounting, Auditing & Taxation*, Vol. 9 No. 2, pp. 159-83.

Beattie, V. and Jones, M.J. (2001), "A six-country comparison of the use of graphs in annual reports", *International Journal of Accounting*, Vol. 36, pp. 195-222.

Birnberg, J.G., Turopolec, L. and Young, S.M. (1983), "The organizational context of accounting", *Accounting, Organizations and Society*, Vol. 8 No. 2/3, pp. 111-29.

Cleveland, W.S. and McGill, R. (1987), "Graphical perception: the visual decoding of quantitative information on graphical displays of data", *Journal of the Royal Statistical Society A*, Vol. 150 No. 3, pp. 192-229.

Courtis, J.K. (1997), "Corporate annual report graphical communication in Hong Kong: effective or misleading?", *The Journal of Business Communication*, Vol. 34 No. 3, pp. 269-88.

DTI (2000), *Modern Company Law for a Competitive Economy: Final Report*, July, Company Law Review Steering Group, London.

FRC (1995), *FRC Progress Report*, Financial Reporting Council, London.

FRC (1999), *Annual Review 1998*, Financial Reporting Council, London.

Frownfelter, C.A. and Fulkerson, C.L. (1998), "Linking the incidence and quality of graphics in annual reports to corporate performance: an international comparison", *Advances in Accounting Information Systems*, Vol. 6, pp. 129-51.

GRI (2000), *Sustainability Reporting Guidelines on Economic, Environmental, and Social Performance*, June, Global Reporting Initiative, Boston, MA.

Hanson, J.D. (1989), "Developments in financial reporting over the last 20 years", in Skerratt, L.C.L. and Tonkin, D.J. (Eds), *Financial Reporting 1988-89: A Survey of UK Published Accounts*, Institute of Chartered Accountants in England and Wales, London, pp. 3-13.

Harris, R.L. (1996), *Information Graphics: A Comprehensive Illustrated Reference*, Management Graphics, Atlanta, GA.

Henry, G.T. (1995), *Graphing Data: Techniques for Display and Analysis*, Applied Social Research Methods Series, Vol. 36, Sage Publications, CA.

Hopwood, A.G. (1996), "Editorial", *Accounting, Organizations and Society*, Vol. 21 No. 1, pp. 55-6.

Johnson, J.R., Rice, R.R. and Roemmich, R.A. (1980), "Pictures that lie: the abuse of graphs in annual reports", *Management Accounting*, October, pp. 50-6.

Korol, J.K. (1986), "Graphical perception and the representation of financial information", *Georgia Journal of Accounting*, Spring, pp. 147-57.

Kosslyn, S.M. (1989), "Understanding charts and graphs", *Applied Cognitive Psychology*, Vol. 3, pp. 185-226.

Kosslyn, S.M. (1994), *Elements of Graph Design*, W.H. Freeman, New York, NY.

Lee, T.A. (1994), "The changing form of the corporate annual report", *Accounting Historians Journal*, June, pp. 215-32.

Leivian, G.M. (1980), "How to communicate financial data more efficiently", *Management Accounting (USA)*, July, pp. 31-4.

Levitt, A. (1998), "The numbers game", speech at the NYU Center for Law and Business, New York, 28 September, available at: www.rutgers.edu/Accounting/raw/aaa/newsarc/pr101898.htm (accessed 20 October 1998).

Lewellen, W.G., Park, T. and Ro, B.T. (1996), "Self-serving behavior in managers' discretionary information disclosure decisions", *Journal of Accounting and Economics*, Vol. 21, pp. 227-51.

Libby, R., Bloomfield, R. and Nelson, M.W. (2001), "Experimental research in financial accounting", working paper, Cornell University, Ithaca, NY.

Libby, R. and Lewis, B.L. (1977), "Human information processing research in accounting: the state of the art", *Accounting, Organizations and Society*, Vol. 2, pp. 245-68.

Maines, L.A. and Hand, J.R.M. (1996), "Individuals' perceptions and misperceptions of time series properties of quarterly earnings", *The Accounting Review*, Vol. 71 No. 3, pp. 317-36.

Mather, P., Ramsay, A. and Serry, A. (1996), "The use and representational faithfulness of graphs in annual reports: Australian evidence", *Australian Accounting Review*, Vol. 6 No. 2, pp. 56-63.

Mather, P., Ramsay, A. and Steen, A. (2000), "The use and representational faithfulness of graphs in IPO prospectuses', *Accounting, Auditing & Accountability Journal*, Vol. 13 No. 1, pp. 65-83.

Murphy, K.J. and Zimmerman, J.L. (1993), "Financial performance surrounding CEO turnover", *Journal of Accounting and Economics*, Vol. 18, pp. 273-315.

Preston, A.M., Wright, C. and Young, J.J. (1996), "Imag[in]ing annual reports", *Accounting, Organizations and Society*, Vol. 21 No. 1, pp. 113-37.

Revsine, L. (1991), "The selective financial misrepresentation hypothesis", *Accounting Horizons*, December, pp. 16-27.

Simcox, W.A. (1984), "A method for pragmatic communication in graphic displays", *Human Factors*, August, Vol. 26 No. 4, pp. 483-7.

Squiers, C. (1989), "The corporate year in pictures", in Bolton, R. (Ed.), *Contest of Meaning*, Massachusetts Institute of Technology, Cambridge, MA, pp. 207-18.

Steinbart, P.J. (1989), "The auditor's responsibility for the accuracy of graphs in annual reports: some evidence of the need for additional guidance", *Accounting Horizons*, September, Vol. 3 No. 3, pp. 60-70.

Taylor, B.G. and Anderson, L.K. (1986), "Misleading graphs: guidelines for the accountant", *Journal of Accountancy*, October, pp. 126-35.

Tufte, E.R. (1983), *The Visual Display of Quantitative Information*, Graphic Press, Cheshire, CT.

Tweedie, D. and Whittington, G. (1990), "Financial reporting: current problems and their implications for systematic reform", *Accounting and Business Research*, Winter, Vol. 21, pp. 87-102.

Wainer, H. (1992), "Understanding graphs and tables", *Educational Researcher*, Vol. 21 No. 1, pp. 14-23.

Watts, R. and Zimmerman, J.L. (1986), *Positive Accounting Theory*, Prentice-Hall, Englewood Cliffs, NJ.